Developing a Pool of Test-items to Assess Student Understanding of the History of Science : Based upon AAAS "Benchmarks for Science Literacy."

Sun-Young Ko*

Seoyun middle school, Seodaemun-gu, Seoul

Seung-Un Choe

Department of Earth Science Education, Seoul National University

ABSTRACT

The nature of science, history of science and philosophy of science have become more and more important in modern society, with an emphasis on 'science literacy' and 'science for all'. In this study, a pool of test-items has been developed in Korean to assess student understanding of the history of science.

The Items were based on chapter 10, "Historical Perspective", in "Benchmarks for Science Literacy" (1993) published by AAAS. Its target grades are from 6th to 12th and the pool consists of 576 test-items. Reliability and validity were tested and item analysis was performed.

The developed test-items can be customized by reconstruction or modification according to the situation, and are able to contribute to science education for establishing science literacy.

Key words : History of science, historical perspective,

* Corresponding author : avior@chol.com

Benchmarks for Science Literacy, Nature of science, Science literacy

I. Introduction

In modern society, science literacy is essential to everybody. The object of modern science education does not give weight to training scientists or scientific experts anymore. It is more important for students to recognize and understand the importance of scientific ideas and evidence and to acquire the ability to solve problems. Most science educational experts and researchers agree that 'scientific literacy' is based on an understanding of the nature of science rather than scientific knowledge (McComas, 1998; Driver et al., 1996; Smith & Scharmann, 1999). Many results of research also show that 'scientific literacy' or 'the nature of science' can be taught efficiently through learning the history of science (Argyll, 1918; Mattews, 1992; Song, S.-Y., 1971; Yang, S.-H., 1996; Lee, S.-G., 1994; Solomon, 1992; Lockhead & Dufresne, 1989; Lee, G.-Y., 1998). Thus, it is necessary to assess student understanding of the history of science if they are to fulfill their scientific roles and a kind of tool for evaluation and/or assessment should be developed.

Therefore, the goal of this research is to develop a tool consisting of items which can assess 6~12th grade student understanding of the history of science, which will be based on "The Benchmarks for Science Literacy; chapter 10. Historical Perspective," published by AAAS (American Association of Advancement for Science) as a part of the American 'Project 2061' for science literacy. The contents in "Benchmarks" are applicable for current

science classrooms in Korea as a reference for the goals of science education.

II. Procedure of developing test items

A. An outline

A previous study, "Development of a Pool for Test-Items to Assess Students' Progress in an Understanding of the Nature of Science." (Lee, E.-A., 2001), proceeded and helped to develop test-items assessing student understanding of the history of science. To develop the test items, we made a set of procedures as shown in Figure 1. We followed the procedure, which was used to develop test items for assessing the Nature of Science (Lee, E.-A., 2001). First of all, we first set the purpose of assessment and figured out what students should understand about the history of science. Then, we carefully examined the benchmark and selected the sentences adapted to the purpose we set. Next we had to decide how the test items should be given to students in the assessment strategy stage and started to develop the items according to the strategy. After finishing the temporary test items, we checked reliability and validity of the items by reviewing them and rewrote the test items based on the reviews.

Roughly, the developmental steps were taken over thirteen months from July, 2002 to August, 2003 and the review took place over four months from September to December, 2003.

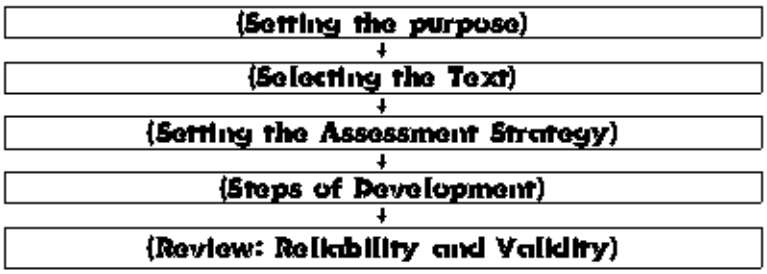

Figure 1. Procedure of test item development (adopted by Sun Ah Lee, 2001)

B. Benchmarks For Science Literacy

The Benchmarks For Science Literacy is the second report of Project 2061, presenting the goals of school science achievement for each of the grade spans from k-2, 3-5, 6-8, and 9-12. Chapter 10, Historical Perspectives in Benchmarks for Science Literacy, states the standards connected to selected episodes related to the history of science, which were based on developing the test items.

C. A strategy of the developed test items

The items in this research have been developed according to criterion-referenced evaluation. Even though their difficulty was not merely considered but also corresponded with the standards set out in Benchmarks, the test items were developed as a form of 'True or False', which has some weaknesses such as the students' ability to guess answers, but also has many merits in terms of reliability, validity, objectivity and practicality. Above all, it is a very useful method because of its wide application to educational fields.

The reviewed items were developed as an item pool, consisting of types I, II and III, each of which covers the same content in the history of science (see appendix). The main purpose of developing this kind of pool is to prevent students' guessing and to reduce bias in interpreting and/or analyzing students' results after assessing them.

D. Steps of item development

According to the outline, we followed 8 steps for developing each item as shown in Figure 2. After these steps, we obtained 3 kinds of true-false question items, covering the same content in the history of science.

Step1	**(Benchmarks statement translation)**
Step2	**(Selecting key sentences)**
Step3	**(Transforming the sentence to true-item)**
Step4	**(Producing the false-item)**
Step5	**(Searching for the related sentence)**
Step6	**(Transforming the sentence to a true-false item)**
Step7	**(Producing the new sentence)**
Step8	**(Transforming the sentence to a true-false item)**

Figure 2. 8 steps of item development

E. Serial number of the items

All of the items in this research were given serial numbers to identify them. Each number contains each item's unique information such as grade level, sub-scale, item type and Benchmarks statement in which the item was referred. The number gives five pieces of information as shown in the example in Figure 3.

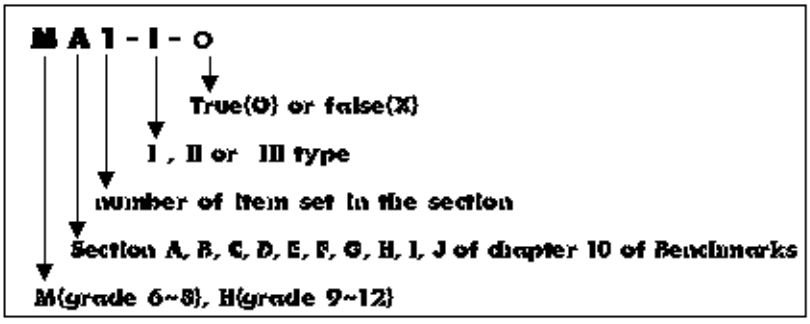

Figure 3. Serial number of each item shows its own information.

III. Review of the developed test items

In order to judge the quality of the developed items, reliability and validity were tested, and item analysis was performed on each item.

A. Reliability

The reliability was tested with Cronbach α, using 80 second grade students from a middle school in Seoul to test the items made for 6~8 grades assessment in the history of science, and 72 university students attending the earth science class in Seoul National University were used to assess 9~12 grade items in the history of science. The mean of coefficient of reliability was 0.8867 for the 6~8 grades and 0.9128 for the 9~12 grades.

B. Review of validity

1. Content validity

The test items-pool developed in this research is based on the Benchmarks. Therefore if the contents of the Benchmarks are reflected well in these items, then it can be

regarded that content validity of the Benchmarks supports the content validity of these items. The content validity of the Benchmarks is already authorized, so verification is necessary to show whether the items reflect well the contents of the Benchmarks.

2~3 experts in physics, chemistry, biology and earth science, evaluated each test item by asking the question, "Does each item reflect the concept and key-idea of the Benchmarks efficiently?" According to their response to the above question, the result was that each item reflects the key-idea of the Benchmarks efficiently.

2. Construct validity

The construct validity was tested to obtain a correlation coefficient. The correlation coefficients derived between the total marks and the mark of each I, II and III type were 0.918 0.820 0.941 for the 6~8 grades and 0.909 0.895 0.910 for the 9~12 grades respectively. The correlation coefficients between the total marks and the mark of each section A~J were 0.663 0.903 0.676 0.869 0.770 (A. F. G. I and J) for the 6~8 grades and 0.890 0.690 0.614 0.703 0.673 0.662 0.741 0.687 0.464 (A, B, C, D, E, F, G, H and J) for the 9~12 grades respectively. Thus it is clear that there is very high correlation between the total marks and the mark of each element of the developed test items. The three types were well constructed based upon the same content of the Benchmarks because the correlation coefficients were obtained to be 0.536 0.660 0.761 among the three types respectively.

C. Analysis of the items

All of the developed items are divided into the 4

science classes of physics, chemistry, biology and earth science and 2~3 experts of each major carried out item analysis.

The experts involved in the item analysis were 9 teachers of middle and high school, and 4 of them were on masters or Ph. D courses.

Eight questions made up a questionnaire for each item in the item analysis. And the answer of each question was analysed on a Likert -scale. Averages of scores were more than 4 out of 5 in most cases. But, the items below 3 on average were modified after discussion with the graders.

In order to verify the property of difficulty of linguistic terms, all items for the 6~8 grade were tested with 5 8th grade students before testing the reliability. We reflected this result to confirm the final items.

IV. Research Results

According to the procedure, the test-items assessing student understanding of the history of science based on AAAS "Benchmarks for Science Literacy (AAAS, 1993)" have been developed and formed to the test item-pool. The quality of the developed items was tested by checking the reliability, validity and item analysis.

Each set of items dealing with the same content in the history of science is organized with 6 question items. All of the question items are developed and constructed with 576 question items in 288 pairs of 96 sets. Table 1 shows a construction of one item set. Table 2 shows the construction of the whole set of items. A small part of the pool is shown as an example in the appendix.

Table 1. Construction of 1 set

Dealing with the same content in the history of science	Type I	True ○	One set (six items)
		False ×	
	Type II	True ○	
		False ×	
	type III	True ○	
		False ×	

Table 2. Constitution of whole items

SECTION	Grade 6-8	Grade9 -12	Subtotal
A DISPLACING THE EARTH FROM THE CENTER OF THE UNIVERSE - Copernican revolution	5 Sets (30 items)	12 Sets (72 items)	17 Sets (102 items)
B UNITING THE HEAVENS AND EARTH - Newtonian synthesis	□	9 Sets (54 items)	9 Sets (54 items)
C RELATING MATTER & ENERGY AND TIME & SPACE - relativity theory	□	11 Sets (66 items)	11 Sets (66 items)
D EXTENDING TIME - the age of the earth	□	5 Sets (30 items)	5 Sets (30 items)
E MOVING THE CONTINENTS - plate tectonics	□	5 Sets (30 items)	5 Sets (30 items)
F UNDERSTANDING FIRE - Lavoisier and controversy over the nature of burning	8 Sets (48 items)	5 Sets (30 items)	13 Sets (78 items)
G SPLITTING THE ATOM - discovery of radioactivity and the structure of the nucleus	3 Sets (18 items)	8 Sets (48 items)	11 Sets (66 items)
H EXPLAINING THE DIVERSITY OF LIFE - evolution by natural selection	□	10 Sets (60 items)	10 Sets (60 items)
I DISCOVERING GERMS - development of germ theory	8 Sets (48 items)	□	8 Sets (48 items)
J HARNESSING POWER - Industrial Revolution	4 Sets (24 items)	3 Sets (18 items)	7 Sets (42 items)
Subtotal	28 Sets (168 items)	68 Sets (408 items)	96 Sets (576 items)
Total	96 Sets(576 items)		

As an example of applying the test items to students, we want to present the additional results of analyzing data for testing the reliability. As mentioned

above, the target students were 80 middle school students in Seoul and 72 university students at Seoul National University. Figure 4 and Figure 5 show how well each group of testing students understand each content categories related to the history of science The test results show high understanding in earth science . The result is reasonable and can be explained as the effect of the history of science being especially emphasized in the earth science curriculum of secondary education. Figure 6 shows university students' understanding of the history of science in terms of grade. It shows that students' understanding the history of science tends to go up with the grade.

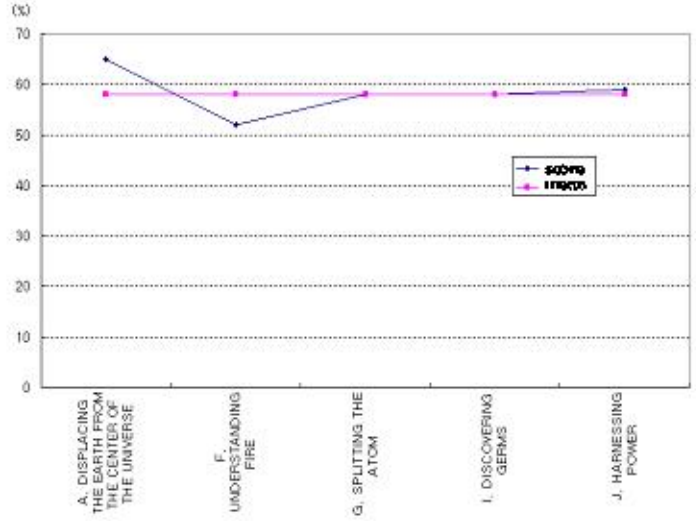

Figure 4. Understanding of the history of science with respect to content categories - middle school students

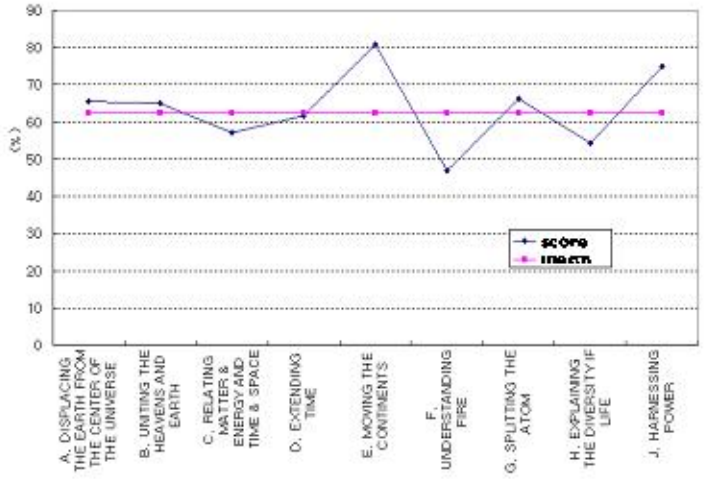

Figure 5. Understanding of the history of science with respect to content categories - university students

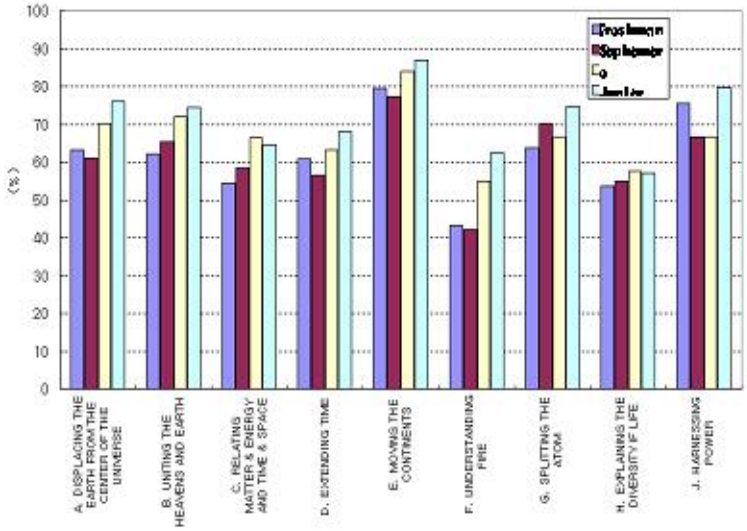

Figure 6. Understanding of science history with respect to grade - university student

V. Discussion and Application

The developed test items can be customized by reconstructing and modifying them to fit the purpose of each test and the level of the target students. The reliability and validity should be confirmed on each test. Both in terms of reducing the possibility of guessing and improving the accuracy of analysis, the examiner can write additional responses such as 'seems correct', 'seems wrong', 'I don't know', 'I can't understand the question' in the 'correct' and 'wrong' response sheet. Furthermore, a pair of true or false items can be used at the same time to reduce the possibility of guessing and miscoding by a confusion of grade and name of examinee. The analysis and interpretation can be variously changed according to the purpose of testing: what we want to know about related to the history of science.

The developed items give a criterion-referenced evaluation rather than norm-referenced evaluation. Therefore, the result analysis should focus on the students achieving the object of criterion rather than the division of superiority and inferiority by relative order. The result of this test can give meaningful information not only as evaluating student understanding (or learning) but also as evaluating teaching itself in the science education curriculum. Therefore the result of this test should be reflected in increasing teaching methods and learning abilities in the history of science of the given science education curriculum as well as deciding on the object and process of science education in the near future.

VI. Conclusion

Project 2061 is one of the reform movements within science education aimed at cultivating 'science literacy' and 'science for all people' after the crisis of science in 1980's. Therefore in this research the pool consisting of 576 test items to assess student's understanding of the history of science for the 6~12 grades was developed in Korean, based on "the Benchmarks for Science Literacy; chapter 10. Historical Perspective" published by AAAS (American Association of Advancement for Science) as a part of the American 'Project 2061' for science literacy.

The quality of the developed items was investigated through reliability, validity and item analysis. Therefore we can expect that the developed test item-pool can help the cultivation of science literacy for students.

The developed items give a criterion-referenced evaluation through which we can easily customize the items by modifying and improving them to fit the purpose of each test and the level of the target students.

Therefore the result of this kind of test should be helpful for improving teaching methods and student's learning abilities in the history of science within the given science education curriculum as well as deciding the object of science education in the near future.

References

An academy of Korea science education (1998). *A wordbook of science education.* Seoul: Educational science.

American Association for the Advancement of Science (1989). *SCIENCE for All Americans.* New York: Oxford University Press.

American Association for the Advancement of Science (1993). *BENCHMARKS for Science Literacy.* New York: Oxford University Press.

American Association for the Advancement of Science (1998). *BLUPRINTS for Reform.* New York: Oxford University Press.

Halliday, Resnick, Walker (1993). *Fundamentals of physics : extended, whit modern physics.* Canada: Jone wiley & sons, Inc.

Hellman. H., (1999). *Great feuds in science.* John Wiley & Sons, Inc.

Hodson, D., (1998). *Teaching and Learning Science.* Buckingham Philadelphia : Open University Press.

Kim, M.-R., (2000). *The effect of science history about 'Electricity and Magnetism' on students' attitude science and orientations to the nature of science.* A master's thesis of Seoul national University.

Kim. J.-S., Lee Y.-D., Hwang, G., Lee, H.-W., (1994). *A curriculum and an evaluation.* Seoul: Educational science.

Kim, Y-S., Park, S.-R., Song, S.-Y., Sung, Y.-G., (1992). *History of Science.* The Ministry of Education.

Lee, E.-A., (2001). *Development of a Pool for Test-Items to Assess Students' Progress in an Understanding of the Nature of Science.* A doctor's thesis of Seoul national University.

Lee, K.-Y., (1998). *A study on the instrution of earth science*

concept-learning by using the history of science. Seoul national University. Master's thesis.

Leite, L., (2002). History of Science in Science Education: Development and Validation of a checklist for Analysing the Historical Content of Science Textbooks. *Science & Education*, 11( ), 333-359.

Lim, K.-S., (1996). *Issues in science of 20th century.* Seoul: Minumsa.

Matthews, M. R., (1994). *Science teaching : the role of history and philosophy of science.* New York : Routledge

Mink, M., Osborne, J., (1997). Placing the history and philosophy of science curriculum: A model for the development of pedagogy. *Science Education* 81 405-242.

National Research Council. (1996). *National Science Education Standards*. Washington DC: Academy press.

Osborne. J., (1998). Learning and Teaching about the Nature of Science. In Mary Ratcliffe (Ed.), *ASE Guide to Secondary Science Education*. Stanley Thornes.

Park, S.-J., Jo, H.-Y., (1998). *The research of science education.* Seoul: Educational science.

Park, W.-H.,, Jang, J.-S., Ha, M.-K., (2002). *The Method of educational research and analysis of statistics.* Seoul: Yangseowon.

Ratcliffe, M., (1998). The Purposes of Science Education. In Mary Ratcliffe (Ed.), *ASE Guide to Secondary Science Education*. Stanley Thornes.

Reiss M. J., (1998). Science for All. In Mary Ratcliffe (Ed.), *ASE Guide to Secondary Science Education*. Stanley Thornes.

Seo, S.-M.,(1996). *The analysis of high school students' understanding of the nature of science.* A master's thesis of Seoul national University.

Sim, W.-Y., (2001). *The Method of educational research.*

Seoul: Haw.

Solomon, J., (1991). *Exploring the Nature of Science*. Glasgow. Blackie.

Sung, H.-H., (2001). *The effect if teaching conservation of energy through history* .A master's thesis of Seoul national University.

Sung, T.-J., (1998). *Practice of item development and analysis.* Seoul: Hakjisa.

Sung, T.-J., (1999). *Understanding of educational research's methods*. Seoul: Hakjisa.

Sung, T.-J., (2002). *Validity and reliability.* Seoul: Hakjisa.

Yang, S.-h.,, Song, J.-W.,. Kim, I.-W., Jo, J.-I.,, Jung, W.-W., (1996). *The history of science and science education.* Seoul: Deawoohaksulchongseo

Yoo. J.-S., (1997). *The study about elementary students' conceptions toward the nature of science.* A master's thesis of Seoul national University.

Appendix

Benchmarks' original sentence

The motion of an object is always judged with respect to some other object or point and so the idea of absolute motion or rest is misleading.

MA1-I-O

The motion of an object is always judged with respect to some other object or point and so the idea of absolute motion is misleading.

MA1-I-X

The motion of an object is always judged with respect to some other object or point and so the idea of absolute motion is not misleading

MA1-II-O

The motion of an object is always judged with respect to some other object or point and so the idea of absolute rest is misleading.

MA1-II-X

The motion of an object is always judged with respect to some other object or point and so the idea of absolute rest is not misleading.

MA1-III-O

Motion is differently observed as an observer stands still or not.

MA1-III-X

Motion is equally observed regardless of observer's motion.

Benchmarks' original sentence

That perception(the earth is large and stationary and other objects on the sky orbit around it) was the basis for theories of how the universe is organized that prevailed for over 2,000years.

HA2-I-O

Geocentricism which prevailed for over 2,000years before Copernicus was based on people's perception.

HA2-I-X

Geocentricism which prevailed for over 2,000years before Copernicus was not based on people's perception.

HA2-II-O

People's perception is relatively important rather than logical reasoning sometimes in the history of science.

HA2-II-X

People's perception is rarely relatively important rather than logical reasoning in the history of science.

HA2-III-O

Ancient people judged the principles of universe's operation by appearances.

HA2-III-X

Modern people judge the principles of universe's operation by only appearances.